DOCUMENT RESUME

ED 325 831 CS 010 318

AUTHOR Anderson, Richard C.; And Others
TITLE A Microanalysis of the Small-Group Guided Reading Lesson: Effects of an Emphasis on Global Story Meaning. Technical Report No. 519.
INSTITUTION Bolt, Beranek and Newman, Inc., Cambridge, Mass.; Illinois Univ., Urbana. Center for the Study of Reading.
SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE Dec 90
CONTRACT G0087-C1001-90
NOTE 47p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Grade 3; *Instructional Effectiveness; Learning Processes; *Oral Reading; Primary Education; *Reading Comprehension; Reading Improvement; *Reading Instruction; Reading Processes; Reading Research; *Small Group Instruction
IDENTIFIERS Guided Reading Procedure; *Meaningfulness; Story Content

ABSTRACT

A study sought to replicate and extend previous research which found that an emphasis on meaning in reading leads to better recall of lesson material than does an emphasis on accurate oral reading, and that the child who is taking an active turn recalls more of the lesson material than do the children who are following along. Six third-grade classes (149 children) received 2 lessons in which the teaching emphasis was on global story meaning and 2 lessons in which the teaching emphasis was on surface features of language. The study demonstrated that an emphasis on story meaning leads to superior performance on an array of outcome measures, especially for children in low and average reading groups. The study confirmed that children taking an active turn reading aloud and answering the teacher's questions, particularly children whose reading fluency is low, learn more at these moments. The study found that the average ability of a reading group, especially average group fluency, is more strongly related to outcome measures than are the abilities of individual members of the group. The study showed that the importance and density of information on a page and the serial position of the page, but not its readability, are associated with outcomes. (Six figures and 14 tables of data are included, and 52 references are attached.) (Author/SR)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED325831

CENTER FOR THE STUDY OF READING

Technical Report No. 519

A MICROANALYSIS OF THE SMALL-GROUP GUIDED READING LESSON: EFFECTS OF AN EMPHASIS ON GLOBAL STORY MEANING

Richard C. Anderson
Ian A. G. Wilkinson
Jana M. Mason
University of Illinois at Urbana-Champaign

December 1990

University of Illinois at Urbana-Champaign
51 Gerty Drive
Champaign, Illinois 61820

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

D. Anderson

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

CS010318

The work upon which this publication was based was supported in part by the Office of Educational Research and Improvement under Cooperative Agreement No. G0087-C1001-90 with the Reading Research and Education Center. The publication does not necessarily reflect the views of the agency supporting the research.

EDITORIAL ADVISORY BOARD
1990-91

James Armstrong
Gerald Arnold
Diana Beck
Yahaya Bello
Diane Bottomley
Clark A. Chinn
Candace Clark
John Consalvi
Irene-Anna N. Diakidoy
Colleen P. Gilrane
Barbara J. Hancin
Richard Henne
Michael J. Jacobson
Carole Janisch
Bonnie M. Kerr
Paul W. Kerr
Daniel Matthews
Kathy Meyer Reimer
Montserrat Mir
Jane Montes
Juan Moran
Anne Stallman
Bryan Thalhammer
Marty Waggoner
Ian Wilkinson
Hwajin Yi

MANAGING EDITOR
Fran Lehr

MANUSCRIPT PRODUCTION ASSISTANTS
Delores Plowman
Debra Gough

Abstract

Six third-grade classes received two lessons in which the teaching emphasis was on global story meaning and two lessons in which the teaching emphasis was on surface features of language. The study demonstrated that an emphasis on story meaning leads to superior performance on an array of outcome measures, especially for children in low and average reading groups. The study confirmed that children taking an active turn reading aloud and answering the teacher's questions, particularly children whose reading fluency is low, learn more at these moments. The study found that the average ability of a reading group, especially average group fluency, is more strongly related to outcome measures than are the abilities of individual members of the group. The study showed that the importance and density of information on a page and the serial position of the page, but not its readability, are associated with outcomes.

A MICROANALYSIS OF THE SMALL-GROUP, GUIDED READING LESSON: EFFECTS OF AN EMPHASIS ON GLOBAL STORY MEANING

Every day 10 million children throughout the United States gather in small groups for lessons with their teachers in which they read and discuss stories. These small-group, guided reading lessons probably are the most productive part of the reading program. Confirming common sense, teacher effectiveness research suggests that reading growth is greatest when children are directly working with the teacher (for a review, see Rosenshine & Stevens, 1984). In most primary school classrooms, children receive the teacher's most concentrated attention during the small-group lesson and, likewise, children are more attentive during these lessons (Mason, 1984).

Despite the pervasiveness, probable effectiveness, and staying power of the small-group, guided reading lesson in American primary school classrooms, there has been only a modest amount of systematic research on features of this kind of lesson. According to Duffy (1981), one problem is that "reading research and teaching research have been moving forward 'out of earshot' of each other. Research on reading reflects little of what has been discovered about teaching" (p. 113). Another problem is the difficulty and expense of systematic classroom research.

The present study is one step in a program of research that seeks a comprehensive understanding of the small-group, guided reading lesson. The purpose of this research is to determine the factors that converge at a moment to determine whether information then being covered in a small-group reading lesson will be comprehended, learned, and later remembered. The operative phrase is "converge at a moment"; it is for this reason that we call the approach microanalysis. Investigated are the page-by-page and even proposition-by-proposition transactions during the lesson, and how these relate to children's later ability to retell the story and answer questions.

According to Au and Mason (1981), microanalysis of lessons requires *proximal indices*--that is, measures that arise directly from students' performance during the lessons. An advantage of microanalysis using proximal indices is that it gives a chance of untangling the nexus of causal forces that operate moment by moment during a lesson. This is less true of research that uses measures distantly related to classroom transactions--for instance, a standardized end-of-year reading test or an especially prepared end-of-unit test, of say, ability to draw inferences from a new text. On the other hand, microanalysis using proximal indices has drawbacks as well. One disadvantage is that a proximal index does not assess maintenance or transfer of abilities that may have been enhanced by a lesson. In general, there are trade-offs between the close-up view provided by microanalysis and the wide-angle view provided by most other instructional research.

The research described in this report sought to replicate and extend the studies reported by Anderson, Mason, and Shirey (1984). Two of the major findings from this research were that an emphasis on meaning leads to better recall of lesson material than does an emphasis on accurate oral reading, and that the child who is taking an active turn recalls more of the lesson material than do the children who are following along. It was these two findings that the research described herein attempted to replicate. The difference is that this research used grade-appropriate stories, the children met in their regular reading groups, and they were taught by their regular teacher; whereas the previous studies employed lists of unrelated sentences, the children met in especially constituted groups, and they were taught by a research assistant.

The first major issue investigated in the present study was whether an emphasis on meaning leads to better student performance than does an emphasis on word analysis and accurate oral reading. An emphasis on meaning was achieved by having teachers ask questions that required students to predict upcoming story events. Under the comparison condition, teachers maintained an emphasis on surface

features of language by asking questions that called for word analysis and by stressing accurate oral reading. As will be explained in more detail later, the theory is that predicting upcoming story events leads children to construct interconnected representations of the stories, which is expected to have benefits for comprehension, learning, and remembering.

The second major issue investigated in the study is whether a child who is the *turntaker* at any given moment--that is, the one reading aloud, answering a question, or making a contribution to a discussion--is getting more from the lesson at this moment than *nonturntakers*--that is, children who are following along. The theory to explain the presumed benefits of active turntaking is simple: Being the turntaker heightens a child's attention and involvement. Everyone believes that active turntaking is important. Aside from Anderson, Mason, and Shirey (1984), however, we do not know of any direct evidence clearly establishing its importance.

Review of Previous Research

To place the present study in context, we will summarize two lines of previous research. The first is research on the relative effects of high-level and low-level questions on children's learning, which is relevant for understanding the probable effects of the types of questions employed in this study. The second is research on grouping children for instruction, which is important because of the possibility that children in low, average, and high reading groups will respond differently to lessons that emphasize story meaning, lessons that emphasize the surface features of language, and to turntaking opportunities.

Teacher Questions and Learning

Questions are perhaps the major tool a teacher has for achieving a certain emphasis in a lesson. In a classic study reported in 1967, Guszak found that nearly 80% of the questions asked by a sample of second-, fourth-, and sixth-grade teachers were low-level, literal questions whereas only 20% could be classified as high-level, nonliteral questions. Until recently, there has been little reason to suppose that teachers have not continued to ask mainly literal questions. For instance, literal questions have predominated in basal manuals (Hansen, 1981). However, in 1989 O'Flahavan, Hartman, and Pearson (1989) repeated Guszak's study and found a sizable shift in the kinds of questions teachers are asking. According to their data, 50% of the questions teachers now ask are literal and 50% are nonliteral.

Evidence that high-level questions actually promote the reading of primary grade children is neither plentiful nor completely persuasive. Tierney and Cunningham (1984) conclude that, "Despite the fact that informal prequestioning and discussion are widespread classroom practices, we could find little research which examined the effect of these activities as actually carried out in classrooms on students' learning from reading" (pp. 619-620). And later they state, "Too few studies have examined the effectiveness of using questions within classroom settings" (p. 623). In seeming contradiction to Tierney and Cunningham's conclusion is Redfield and Rousseau's (1981) meta-analysis of classroom questioning studies. They found an advantage for higher order questions. But none of the titles of the 14 studies that could be included in the analysis mentions reading; many mention social studies or science instead. Only a few studies involved primary or elementary school students.

In fact, in reading in the primary grades there is some evidence that low-level questions might even lead to greater gains in reading achievement than high-level questions (for a review see Rosenshine & Stevens, 1984). In a reanalysis of the large data set from the First Grade Reading Group Study, Martin, Veldman, and Anderson (1980) found that answering "reading" questions ("questions that required students to decode words and focus on sounds" p. 487) and answering what were infelicitously called "nonreading" questions ("comprehension questions, thought questions, and fact questions") were both positively related to year-to-year gains in reading achievement. Maybe the content of questions is not so important as opportunities for active turns. Or, perhaps low-level questions lead to a larger number

of opportunities for turns and the turns are more evenly distributed across members of the group, balancing what would otherwise be an advantage for high-level questions.

The inconsistent results obtained in previous studies of high-level questions with primary grade children may be attributable to children not knowing how to answer them. Raphael and Wonnacott (1985) taught children how to analyze question-answer relationships and how to draw the inferences that different types of questions require. This instruction, along with lessons in which children got a diet of questions demanding inferences, has produced positive results in classroom investigations.

Many studies have evaluated complicated treatments in which high-level questions are only one feature. For instance, evaluations of reciprocal teaching (e. g., Palincsar & Brown, 1984) and the Directed Reading Thinking Activity (e. g., Davidson, 1970; Petre, 1970, both cited in Tierney & Cunningham, 1984) have been positive. However, it is impossible to say what role the nature of the questions played in these results.

Despite the confusing picture from previous classroom research, we believe that there are compelling theoretical reasons to suppose that asking questions that invite students to predict upcoming story events will enhance comprehension of and memory for stories (Anderson, Mason, & Shirey, 1984). The reasoning is that the task of predicting story events is likely to cause children to construct interconnected story representations. This is important because, phenomenologically, understanding a story is "seeing" how story events are connected. Likewise, retrieving information from memory about any given story event involves having a story representation in which temporal and causal connections with the rest of the story have been established (van den Broek, 1990). In contrast, an emphasis on word analysis and accurate reading is unlikely to promote the formation of interconnected story representations. Indeed, asking word-analysis questions between story segments may actually disrupt children's attempts to make bridging inferences. Similarly, focusing on word-level errors during reading turns may distract children from the task of building overarching connections among story events.

Grouping for Instruction

Grouping has been studied from a number of angles. Much of the previous research has addressed the issue of whether students of a given level of ability achieve more in homogenous or heterogeneous groups. Reviews of this research conclude that being in a homogeneous group sometimes helps high-ability students, but that it fairly consistently retards the achievement of low-ability students and may impair their social status and self-esteem as well (Persell, 1977; Good & Marshall, 1984; though see Slavin, 1987).

A plausible reason for the poor showing of low groups is that they tend to receive less satisfactory instruction (Hiebert, 1983). For instance, the pace of instruction is slower (Barr & Dreeben, 1983), and teachers provide less sustaining feedback in low groups (Anderson, Evertson, & Brophy, 1979; Hoffman et al., 1984). Concomitantly, children in low groups are less attentive than children in high groups (Felmlee, Eder, & Tsui, 1985) and less of the talk in low groups is about the task at hand (Johnson, 1981).

Other research on grouping has taken an organizational perspective. Grouping is seen as a means for coping with the number and diversity of children in a class in a way that enables orderly coverage of the curriculum (Barr & Dreeben, 1983; Dreeben, 1984; Dreeben & Barr, 1987). Among the interesting findings from this type of research is that high groups are larger in racially mixed classes, which gives black students an increased chance to get into high groups (Sorensen & Hallinan, 1984).

Despite the distressing association between membership in low groups and less effective instructional practices, authorities on grouping such as Dreeben (1984) and Hallinan (1984) stress that the

composition of groups and the nature of instruction ought to be conceived as distinct, although interacting factors. In Hallinan's words,

> Grouping practices and mode of instruction may vary separately or jointly across tracks or group levels. . . Ultimately, it should be possible to determine which particular grouping practices are most effective at each ability level and how their effectiveness relates to the mode of instruction that distinguishes that level. Thus, research models should specify how grouping practices and modes of instruction interact to enhance or hinder learning. (pp. 234-235)

To recapitulate, the focal questions posed in the present study were as follows: (a) Which is more effective in a small-group reading lesson, an emphasis on story meaning or an emphasis on word analysis and accurate oral reading? and (b) Do children get more from reading lessons at moments when they are taking active turns?

If one grants that classroom instruction is, or could be, a complex phenomenon involving the interaction of many factors, then it follows that dependable answers to the two foregoing questions cannot be obtained outside of an investigation of other features of small-group instruction that are of known or possible importance. Thus, a number of additional factors were examined in this study. Briefly, these factors included the children's individual reading comprehension level and reading fluency, gender and ethnicity, the average comprehension level and fluency of reading groups, the positions of the pages in the stories the children read, and the importance, readability, and density of the information on each page. Of special concern was whether active turntaking or an emphasis on meaning or surface interacted with the abilities of the children or the average abilities of the reading groups of which they were members.

Method

Subjects

One hundred forty-nine children in six third-grade classes (including one combination second-third grade and one combination third-fourth grade class) from two schools participated in the study. One school was located in a working-class neighborhood and the other in a middle-class neighborhood of a midwestern industrial city. About 75% of the children were white and 25% were black, with a few other minorities. A standardized reading achievement test given during the study showed that subjects had an average stanine of 6.6 with a standard deviation of 2, which compares with a national average of 5 and standard deviation of 2.

Design and Procedure

Six teachers agreed to use four stories with all the children in their classrooms. The stories were featured in lessons taught on Monday, Tuesday, Thursday, and Friday during a week in the spring. A battery of tests was given on Wednesday. Each story, which was completely read in one lesson, was introduced by the teacher with the following instructions:

> I have a story for you to read today. You'll take turns reading this story out loud. This means that when one person is reading out loud, the rest of you should be following along in your own booklets reading silently. Please don't say anything or make any noise when someone is reading because we want to make sure that we hear the whole story. After you read each page of the story I'll ask you a question about the story. Any questions?

Lessons for two stories engaged children in story meaning using prediction questions that focused on the major plot elements of the stories. This was called the *meaning-emphasis* treatment. Lessons for two other stories engaged the children in word analysis and accurate reading. This was called the *surface-emphasis* treatment.

The study employed the partially counterbalanced design displayed in Table 1. Note that receiving the meaning-emphasis (ME) treatment or the surface-emphasis (SE) treatment is orthogonal to both Day (1 through 5) and Story (A through D). Day and Story, however, are partially confounded; this fact is inconsequential because neither Day nor Story holds any intrinsic interest.

[Insert Table 1 about here.]

Teachers taught the lessons to their regular, intact reading groups, all of which were grouped by ability into three reading groups. At the beginning of each reading lesson, teachers distributed booklets containing the day's story. The booklets were in the children's hands only during the lesson. The teachers had children take turns orally reading each page of the story in a fixed order around the reading circle. To assure that order of turntaking was independent of other characteristics of the children, the first child to read in a group was chosen from a predetermined random list.

The teacher recorded the time the lesson began and ended and wrote the name of the child who read each page aloud in her copy of the story booklet. The teacher also marked any reading errors in the booklet, which she returned to us at the end of the lesson. After a child read a page, he or she was asked one question, either about story meaning or word analysis. Questions and likely answers were listed in the teacher's booklet. One volunteer was also called upon to answer; however, only the oral reader's participation was coded as an active turn for the data analysis.

During meaning-emphasis lessons, teachers were advised to "ignore reading mistakes unless the word reported destroys the intended meaning of the sentence. Then correct the error by saying the correct word. Usually repeat the word in a phrase context." During surface-emphasis lessons, teachers were told to "promptly correct every reading error that the oral reader makes. Then have him or her repeat the word. Even if the child only stumbles but self corrects have him or her repeat the word."

Following the story reading, children went to their seats to complete on their own two tasks involving the story they had just read. The tasks were monitored by one of the researchers. First, the children were given a lined paper that began, "Dear friend, Today we read a new story in school. Here is what happened in the story." They were instructed to "Pretend that you are writing a letter to a good friend. You want to tell your friend everything that happened in the story that you just read. Write down everything that happened in the story so that your friend will know what the story is about. We have started the letter for you." One of the members of the research team circulated around the room encouraging children to keep writing.

When children had completed the letter, they were given a sheet containing 12 short-answer questions (13 for one story), one for each page of the story, to answer in writing. For example, the sheet for the story called *Flower Street* included these questions:

> What time of year did the story take place?
> Why did Mrs. Black want to plant seeds?
> How did Mr. and Mrs. Black take care of the plants?
> What did Mr. and Mrs. Black do with the flowers after they were all grown?

Materials

The four stories were selected from children's magazines and books and adapted for use in the study. *The Boy and the Whistle* is a magical story about a child who is led home on a very cold day by the sound of his whistle. *A Groundhog by the Fireplace* describes the misadventures of a foolish man who is deceived into thinking that a groundhog will bring him good luck. *Flower Street* is about the efforts of a woman and her husband to beautify the rundown street they live on by planting flowers. *Instant Watermelon* describes a gullible boy who thinks his father can make a watermelon grow instantly.

All of the reading groups in each classroom read each of the foregoing stories. This was done to enable an assessment of the effect of individual abilities and the average abilities of reading groups that was not confounded by differences in the stories that groups read. We made sure the stories were fairly easy for third graders, so that low groups would not find them too difficult. The average readability of the stories according to the Fry (1977) formula was late second grade (2.8 in grade equivalents). That the stories were, in fact, not too difficult is suggested by the low rate of oral reading errors (see Table 1).

Each story was broken into 12 segments (one story was broken into 13 segments), each segment appearing on a separate page. Segments typically contained 4 or 5 sentences, each beginning with a new paragraph, and presenting a relatively complete idea or event from the story.

Two teacher booklets were constructed for each story, one for the meaning-emphasis condition and one for the surface-emphasis condition. The booklets contained the directions for conducting the lessons and the questions for the teacher to ask. Meaning-emphasis questions asked children to predict what might happen later in the story. For example, the teacher's booklet for *Flower Street* included these questions:

> What might Mr. and Mrs. Black do now?
> What might happen next?
> What might the Blacks do with the plants now?
> What might the people on Flower Street do with the flowers?
> What do you think might happen, now that there are flowers on Flower Street?

The questions asked under the surface-emphasis condition focused children on analysis of words in the section they had just read. For example,the questions asked about *Flower Street* included the following:

> Ask pupils to find two words in this section with a long vowel and silent *e* at the end (*place, space, name*) and one word ending in *e* that does not follow the rule (*there*).
>
> Ask pupils to find words that indicate more than one (plural) (*things, seeds, pots, bowls, cups, jars*).
>
> Point out that flowers has two syllables. Ask them to locate other words that have two syllables (*Stevens, window*).
>
> Explain with the word *Everyone* that it contains two little words and is called a compound word. Ask them to find another compound word (*outside*).

Measures

Comprehension measures were based on subjects' raw scores on the reading subtest from the 1971 edition of the SRA Achievement Series, Primary II, Form E. Raw scores were converted to national stanines using the mean and standard deviation from the testmaker's national norming sample.

Fluency measures were based on the time in tenths of a second that it took subjects to read four lists of 15 words. The lists contained, respectively, high- or low-frequency words with either regular or irregular vowel-sound patterns (Mason, 1976). We summed times on the four lists, took the natural log of the sum to normalize the scores, and inverted the scale before converting to local stanines based on the sample mean and standard deviation.

The measures of recall were based on subjects' written retellings of the stories. Several adults parsed the stories into propositions, where a proposition was approximately equal to an independent clause. An adult rater did blind scoring of the propositions in each retelling. A proposition was counted as recalled if the gist was reproduced, which meant the subject could substitute approximate synonyms, change the syntax, or omit unessential content and still be counted correct. The measure was the percentage of propositions on a page that was recalled. Other adult raters judged the importance of each proposition to the plot of the story. The percentage of these propositions in a story that was recalled constituted the measure of important story elements recalled.

The short-answer measure was based on a lenient scoring of subjects' answers to the questions. A response was counted correct if it matched the response given by any one of several adults who had read the stories and answered the questions, or if it was judged to be a plausible, story-related response.

Oral reading errors were recorded by the teachers during each lesson. An error was defined as any deviation from the print. Teachers were instructed to "put a slash through each word in the story that the student reads incorrectly." One caution regarding the measure of oral reading errors: The teachers were asked to respond to errors differently during meaning-emphasis lessons and surface-emphasis lessons. This could have differentially influenced the criterion they applied to decide whether an error had occurred or whether to record it.

At the end of the week, subjects were presented with a list of the titles of the four stories and asked to rank them in terms of interestingness, where 1 = most interesting and 4 = least interesting. The rankings were later inverted so that a high score meant high interest. It is worth noting that current research suggests that a simple ranking is the most sensitive of several possible measures of children's degree of interest in stories (Jose & Brewer, in press).

Table 2 summarizes basic, descriptive information about the variables in the study. Some of the variables are derived from other variables. For instance, group fluency and fluency of the turntaker of the moment are derivative variables, because each is based on the same scores as individual fluency. Not listed are interactions and the quadratics of continuous variables. Table 2 presents the means and standard deviations of untransformed variables. We applied normalizing transformations to several of the variables before analysis. Percentage of propositions recalled and percentage of important elements recalled were measured in radians following an arcsine transformation ($2 \arcsin\sqrt{p}$). Percentage of oral reading errors was measured in logits ($\ln(p/1\text{-}p)/2$).

[Insert Table 2 about here.]

Data Analysis

Two levels of analysis were completed for each dependent variable. In one, the child was the unit of analysis. In the other, the reading group was the unit of analysis. The latter analysis caters to the theory that a reading group is more than a collection of individuals and deals with the worry that it might not be appropriate to count individual children in a group as contributing independent observations (Burstein, 1980; Raudenbush & Bryk, 1988). Furthermore, in some analyses, called story-level analyses, the basic unit of measurement was performance on a complete story. In other analyses, called page-level analyses, the unit of measurement was performance on one page of text within a story.

Basically, we did analyses of variance. We did the analyses within the flexible logical and computational framework afforded by the general linear model (Cohen & Cohen, 1983; Pedhazur, 1982). Models of most aspects of the data were estimated in ordinary least squares analyses using SPSS[x] Regression (SPSS, 1988). Where the assumptions of ordinary least squares analysis could not be met, we employed maximum likelihood estimation using the SAS Logist procedure from the SUGI supplemental library (SAS Institute, 1986). The nominal alpha level for tests of significance was .01. Hereafter effects will be described simply as significant or nonsignificant.

In analyses employing least squares estimation, tests of significance were constructed using the logic of mixed between-subjects and within-subjects analysis of variance. Between-subjects factors were evaluated in terms of their contribution in explaining between-subjects variance while within-subjects factors were evaluated in terms of their contribution in explaining within-subjects variance. In each case, the error variance was the final residual after all variables had been entered into the equation. This amounts to a pooled estimate of various theoretically possible sources of error. The *F* ratio for each main effect and interaction was computed off-line based on the increment in the squared multiple correlation coefficient at the point of entry.

The degrees of freedom in least squares analyses took account of the number of tests of significance that were run and, of course, whether the analysis was individual level or group level and page level or story level. For example, consider the degrees of freedom in the analysis of the percentage of propositions recalled in which the individual child's performance on an individual page was the unit of analysis. There were 149 children, most of whom recalled propositions from each of 12 pages in each of 4 stories. There were 24 instances when children were absent from lessons; hence, there were 6,864 observations. The degrees of freedom for the error term in within-subjects tests of significance were (N-1)(SP-1)-V = 6,548, where N is the number of subjects, S the number of stories, P the number of pages, and V the number of variables (both main effects and interactions) tested for significance in the full model (120 in this case).

In analyses employing maximum likelihood estimation, the test of significance for a factor was the likelihood ratio test for nested models: twice the difference in log likelihoods of a model including the factor and a model excluding the factor, which is approximately distributed as a chi-square with the degrees of freedom equal to the number of hypotheses tested (usually one). Difference chi-squares are in the same spirit as the point-of-entry *F* ratios employed in least squares analyses.

During the within-subjects phase of analyses, the variables were entered in blocks as follows: within-subjects factors, quadratics of selected within-subjects factors, interactions among within-subjects factors, between-subjects factors, and, finally, interactions between within-subjects and between-subjects factors. The between-subjects factors were evaluated in a separate analysis; they were included in the within-subjects phase of the analysis in order to get proper estimates of interactions of within-subjects and between-subjects factors and appropriate b-weights.

Within blocks of variables, order of entry was based on logic or theory. In cases where there were no a priori grounds, variables were entered with forward selection in order of variance explained. Allow us to stress that the factors of major interest in the study were orthogonal to one another, so in most of the important cases the order in which variables were entered made no difference. Collinearity was an issue only among measures of child characteristics and among measures of text characteristics. In our search for the best model in these cases, we weighed temporal and causal priority and consistency across analyses, as well as the strength of effects in a given analysis. We also compared alternate analyses in which collinear main effects and interactions were entered in different orders.

Here is an example of how we chose among competing models: Group heterogeneity in fluency (the standard deviation of individual fluency scores) competed with group fluency (the mean of individual fluency scores) in almost every analysis. As it happens, there was a pronounced negative correlation

between group heterogeneity and group fluency. That is to say, low groups tended to be composed of children with a wider range of fluency scores than high groups. In most analyses, either group fluency or group heterogeneity, but not both, would make a significant contribution. We chose to include group fluency in the models presented in this paper, instead of group heterogeneity, because group fluency had slightly stronger and more consistent effects, because this made more sense to us, and because previous research suggests that the average level of a group is much more likely to be important than group heterogeneity (Dreeben & Barr, 1987, Table 3, p. 33).

All two-way interactions were investigated, except those among child characteristics and those among text characteristics, because, for instance, the interaction of individual comprehension and group comprehension wouldn't have made any sense. None of the interactions involving three or more factors was examined, because there was no theory to guide the choice of which interactions to examine and unguided exploration would have vitiated the power of the study. Presented here are reduced models that were compiled by deleting factors from the corresponding full models that did not figure in significant effects and rerunning the computer program. Selected nonsignificant effects were included in reduced models to enable comparison with related models in which the effects were significant. Of course, factors involved in significant interactions had to be included in reduced models even when the main effects were not significant.

In page-level analyses, a vector of performance means on the 48 pages (12 pages for each of 4 stories) was entered at the very last step. This removed all remaining variance attributable to page and story and, thereby, increased the precision of the analyses. Unidentified page and story factors had a large, significant effect in every analysis. The b-weights in the tables reflect relationships before the vector of page means was entered.

The percentage of variance accounted for in page-level analyses may seem extremely small. This perception arises because most researchers are accustomed to dealing with performance measures aggregated at least to the level of a story. For the researcher used to dealing with aggregated data, our data have been radically disaggregated in individual-subject, page-level analyses. Although the performance of individual children on single pages of a story is highly variable, which accounts for the small percentage of variance figures, when the performance of dozens of children on dozens of pages is considered, effects can be estimated with approximately the same reliability and precision as with aggregated data. Saying this the other way around, the unexplained or error variance is higher when observations are disaggregated rather than aggregated, but the number of observations is proportionately greater, too, so page-level and story-level analyses can yield roughly the same results.

Results

The first overarching question addressed in this study is whether an emphasis on story meaning or an emphasis on word analysis is more effective in a small-group, guided reading lesson. Table 3 summarizes mean performance as a function of the teaching emphasis on the six dependent variables included in the study. The second major question that we sought to answer is whether turntakers get more from lessons at moments when they are taking turns than do nonturntakers at the same moments. Table 4 presents the mean performance on the two measures included in the study that are relevant to answering this question.

[Insert Tables 3 and 4 about here.]

Next we will present a detailed analysis of each dependent variable. Tables 5 and 6 display alternate analyses of the recall of propositions. Tables 7 and 8 summarize analyses of responses to short-answer questions. Tables 9 and 10 contain analyses of the recall of important story elements. Table 11 shows the analysis of oral reading errors. Tables 12 and 13 present the analyses of children's rankings of

interest in the stories. Finally, Table 14 contains the analysis of lesson time. The figures display the most interesting and important of the interactions.

Table 5 summarizes the analysis in which the dependent variable was the percentage of propositions recalled. The unit of analysis was the individual child's performance on an individual page from one of the stories. The b-weights express the increase in percentage of recall (expressed in logits) attributable to a one-unit change in the variables listed in the left-hand column. The column headed Percent Variance gives estimates of the magnitude of the effects of variables. Each estimate is based on the increment in the squared multiple correlation coefficient at the point at which the variable entered the analysis. In the rows in the table labeled Constant/Residual, the first number is the constant (i.e., the intercept) and the second number is the residual (i.e. the unexplained or error variance).

[Insert Table 5 about here.]

Table 5 establishes that several variables significantly influenced the percentage of propositions recalled. First, the table shows that recall was affected by children's individual comprehension scores and by average group fluency. We had a policy of entering individual measures of ability before the group measures, on the grounds that the individual measure is logically prior. Otherwise, in this and subsequent analyses the group averages usually would have overshadowed the individual measures, because average group ability was more strongly related to outcome measures, even in analyses in which the individual was the unit of analysis.

Second, Table 5 indicates that children recalled more propositions when they read stories under a meaning emphasis than under a surface emphasis. The type of teaching emphasis was involved in one significant interaction: As can be seen in Figure 1, a meaning emphasis produced an advantage over a surface emphasis when average group comprehension was low, but the advantage became smaller as group comprehension increased, nearly disappearing as group comprehension reached a maximum.

[Insert Figure 1 about here.]

Third, Table 5 establishes that children's recall was higher when they were turntakers rather than nonturntakers.

Table 5 indicates that two properties of the information on pages significantly influenced recall of propositions: (a) The more important the information on a page, the greater was the recall and (b) Recall declined from the first page to the last page of the stories. The significant effect of position of page squared appeared because the rate of decline was high at the beginning of a story and leveled off somewhat toward the end.

Table 5 indicates that there were three significant interactions among measures of reading ability and aspects of the information on pages. Figure 2 shows that as average group comprehension increased there was progressively better recall of propositions on pages containing important information whereas group comprehension had a smaller influence on recall of propositions from pages containing less important information. To draw Figure 2, two extreme values of page importance were selected (although the interaction was calculated using all gradations of page importance). The same technique was used to draw the other figures showing interactions of two continuous variables. Therefore, in interpreting interactions involving two continuous variables, it should be kept in mind that the figures show lines representing the upper and lower boundaries of the interactions, enclosing families of unplotted but easily visualizable intermediate lines.

Figure 3 reveals that individual comprehension had a greater influence on recall of propositions on pages early in stories than on pages later in stories. Figure 4 documents that average group fluency had a greater influence on recall of propositions from pages early in stories than later in stories. Although

these two interactions may seem redundant with one another, they were statistically independent. That is, the presence or absence of either interaction in the model did not influence the magnitude or shape of the other.

[Insert Figures 2-4 about here.]

Table 6 summarizes the alternate analysis of percentage of propositions recalled in which the unit of analysis was the performance of an entire reading group on a page of text. Of course, individual-level variables, such as individual comprehension, individual fluency, and turntaking, could not be included in this analysis.

[Insert Table 6 about here.]

A comparison of Table 6 and Table 5 reveals that the analysis based on groups is similar to the analysis based on individual children and only somewhat less sensitive. Most newsworthy is the fact that the meaning-emphasis condition is still significantly superior to the surface-emphasis condition.

Table 7 presents an analysis of the responses to the short-answer questions in which the unit of analysis was an individual child on an individual page. Because there was one question per page, and answers were scored correct (= 1) or incorrect (= 0) with no partial credit, ordinary least squares estimation would have been inappropriate. In its place, we employed logistic regression and maximum likelihood estimation. To provide starting values for the logistic regression, we used ordinary least squares analysis to compute an approximate model. Probabilities based on the approximate model, truncated to lie between 0 and 1, were then expressed as logits and regressed on the reduced model variables, again using least squares regression, giving initial values for the logistic regression.

[Insert Table 7 about here.]

In the within-subjects phase of the analysis of responses to the short-answer questions, we removed all between-subjects variance by entering first a vector of subject means. The b-weights reported in Table 7 reflect this. To control for the effects of unidentified story and pages factors, as we did in other analyses, we removed all remaining variance due to these factors by entering a vector of page means after identified page factors had been entered. The percentages of variance and chi-squares reported in Table 7 reflect this. The vector of page means accounted for 23.76% of within-subjects variance. This is a much larger percentage of variance than is explained by unidentified story and page factors in the individual-level analysis of recall of propositions, because it represents not only story and page variation but also variation in the difficulty of the short-answer questions. The percentages of variance reported in Table 7 use the McKelvey and Zavoina (1975) analog of R^2 for logit models (see Aldrich & Nelson, 1984). These R^2s assume a dependent variable measured on an underlying interval scale, and the sums of squares upon which the R^2s are based are estimates rather than actual values. Hence, the percentage of variance figures should be interpreted cautiously. We report them only to assist comparison with other analyses.

Table 7 shows that several factors and their interactions affected responses to short-answer questions. First, the individual subject's comprehension and fluency and the reading group's average fluency were related to the probability of a correct response.

Second, a meaning emphasis led to a higher probability of correct response than a surface emphasis. The significant interaction of the teaching emphasis with page position is graphed in Figure 5, which indicates that the advantage of a meaning emphasis was greatest early in stories. The comparable interaction was not significant in the analysis in which the dependent variable was percentage of propositions recalled, but the trend was similar. The two functions differ somewhat in shape, but both

show performance tailing off toward the end of a story and both show that the benefits of a meaning emphasis are greatest at the beginning of the story.

[Insert Figures 5 and 6 about here.]

Third, the main effect of turntaking fell short of being significant, but the interaction of turntaking and the turntaker's fluency, shown in Figure 6, was significant: The benefit of an active turn was greater when the turntaker's fluency was low.

As can be seen in Table 7, the density of information on a page, the position of the page in a story, and the quadratic of position significantly affected the probability of correct short-answer responses. The one remaining significant effect was the interaction of page position and group comprehension. High groups did relatively better at the beginning than at the end of stories.

A comparison of Table 7 with Table 5 reveals that in most respects the results with the short-answer measure were similar to the results with the propositions recalled measure.

Table 8 summarizes the alternate analysis of responses to the short-answer questions in which the reading group was the unit of analysis. An ordinary least squares model provided a good fit to these data. The model is comparable to the one obtained in the subject-level analysis presented in Table 7.

[Insert Table 8 about here.]

Table 9 contains an analysis of percentage of recall of important story elements. This is a story-level rather than a page-level variable. Turntaking is a page-level factor so it is not included in this analysis. As can be seen, individual comprehension, individual fluency, group fluency, story, and meaning versus surface emphasis significantly influenced recall of key story elements.

[Insert Table 9 about here.]

Table 10 summarizes the alternate analysis of recall of important story elements in which the reading group was the unit of analysis. The results of this analysis are similar to the results of the one presented in Table 9. In particular, meaning emphasis versus surface emphasis is significant in both analyses.

[Insert Table 10 about here.]

Table 11 contains the analysis of percentage of oral reading errors. The unit of analysis is the child who is reading aloud. This is a page-level variable; in each reading group each page was read aloud by one child.

[Insert Table 11 about here.]

Table 11 indicates that oral reading errors were significantly affected by the turntaker's fluency and comprehension and the density of information on the page. Turntakers made significantly fewer errors when there was a meaning emphasis than when there was a surface emphasis. Two significant interactions appeared. The Meaning vs. Surface x Turntaker's Race interaction appeared because black and white children made the same percentage of oral reading errors when meaning was emphasized, whereas only black children suffered a greater increase in error rate when surface features of language were emphasized. The Meaning vs. Surface x Page Position Squared interaction appeared because, relative to a surface emphasis, a meaning emphasis was associated with more reduction of errors in the middle of a story than at either the beginning or the end.

[Insert Table 12 about here]

Table 12 presents an analysis of the children's rankings of story interest. Each child ranked all four stories on a scale of 1 to 4, so there was no between-subjects variance. Because the scale was ordinal, maximum likelihood estimation was employed in this analysis. We again used the McKelvey and Zavoina (1975) analog of R^2 to estimate the percentage of variance explained. Table 12 documents that an emphasis on meaning caused children to rate the stories as significantly more interesting than an emphasis on surface features of language. Also significant was the particular content and form of the stories. Table 13 contains an alternate analysis of interest, also employing maximum likelihood estimation, using the reading group as the unit of analysis. Again, meaning versus surface emphasis and story were significant.

Table 14 summarizes the analysis of minutes per lesson. This is a story-level variable. Lesson time is an attribute of groups not individuals, thus group was the unit of analysis. Two significant effects were discovered: The higher the average fluency of the group the shorter were the lessons. Meaning-emphasis lessons took less time than surface-emphasis lessons.

[Insert Tables 13-14 about here.]

Discussion

The major finding of this study is the small but pervasive influence of an emphasis on global story meaning. When teachers emphasized meaning by asking questions that required predicting upcoming story events, the children performed somewhat better on every measure included in the study. As compared to the surface-emphasis condition, when meaning was emphasized, the children evidenced more complete recall of propositions, superior recall of important story elements, and more correct answers to short-answer questions; they rated the stories as more interesting, made fewer oral reading errors, and finished the lessons more quickly.

The design of the study suggests that the emphasis on meaning had a causal role in producing these outcomes. This was a counterbalanced within-subjects design; therefore, the outcomes could not have been due to the reading level or other traits of the children, the classroom climate or ability of the teachers, or the characteristics of the stories. The advantage for an emphasis on meaning appeared in both page-level and story-level analyses and whether the individual student or the reading group comprised the unit of analysis.

While the finding that the meaning-emphasis condition was superior to the surface-emphasis condition is notable for its pervasiveness, several issues regarding the robustness and generalizability of the finding must be addressed. A first issue arises from the fact that the teaching emphases examined in this study, especially the surface emphasis, are not representative of most ordinary classroom instruction. The study was designed to determine the limits of the effects of variation in teaching emphasis, and so the treatments were not necessarily intended to be representative. To the extent that the study has been successful, it has established a principle about surface and meaning emphases in small-group reading lessons. However, one must have realistic expectations when generalizing the findings to ordinary instruction. Day in and day out, most teachers probably conduct lessons that lie somewhere in between the surface-emphasis and meaning-emphasis conditions as these were exemplified in this study. Thus, if a typical teacher were more strongly to emphasize story meaning in her lessons, the changes in children's performance probably would not be as pervasive as the ones observed in this study.

The second issue is the small size of the absolute difference between conditions on some measures, especially percentage of propositions recalled and percentage of oral reading errors. One reason the differences seem small is attenuation due to performance floors; the differences probably would have been larger if the average levels of performance had been nearer the midpoints of the scales. The fact that there were performance floors is not a grave problem. It merely makes the absolute size of differences an unsteady guide to the strength of effects. A third, and more specific issue is that there

is no particular reason to suppose that meaning-emphasis instruction would generally take less time than surface-emphasis instruction. It probably took less time in this case because teachers were asked to limit the amount of discussion. The result does indicate that the benefits of meaning-emphasis lessons are not time dependent, a conclusion that also emerged from our previous study (Anderson, Mason, & Shirey, 1984).

That there was actually a higher rate of oral reading errors under the surface-emphasis condition than under the meaning-emphasis condition may seem astonishing, considering that under the former condition accurate oral reading was one of the principal goals toward which teachers were asked to strive. A possible explanation for this finding is that children's developing mental representation for a story provides information that assists decoding (see Pehrsson, 1974). The prediction questions featured in the meaning-emphasis condition presumably promote the development of a coherent story representation, whereas, in contrast, the word-analysis questions asked in the surface-emphasis condition may tend to disrupt the formation of a coherent representation. Furthermore, the stress in the surface-emphasis condition on accuracy while reading may lead some children to make mistakes because they ignore semantic constraints and become overreliant on graphophonemic information. An alternative explanation is procedural: As already indicated, teachers were asked to deal with errors in a different way under the two conditions, and this may have influenced them to apply different criteria for judging whether an error had been made or whether to record it.

Average group comprehension conditioned the effects of an emphasis on meaning on recall of propositions and, to a lesser extent, recall of important story elements. The effects of a meaning emphasis were greatest with low and average reading groups, smaller with high groups. Other research points to the same conclusion. Hansen and Pearson (1982) found that second-grade and fourth-grade students recalled more from stories when preparation for reading included questions designed to activate schemas needed to understand the stories. The advantage was greatest for low readers. Hansen and Pearson surmised that good readers already engage in the processing required for comprehension whereas poor readers often need prompting. The Hansen and Pearson study and the present one converge on the conclusion that poor readers will benefit most from meaning-oriented instruction.

A second major finding of this study was that children learn and remember more lesson material when they are involved in taking active turns. As compared to nonturntakers, children recalled more propositions from pages they had read aloud and answered questions about during the lesson. Taking an active turn also positively influenced responses to short-answer questions when the reading group's fluency was low. Note that propositions recalled and short-answer responses are the only measures included in the study upon which an advantage for active turntaking could have appeared. Note also that turntaking could be evaluated only in subject-level analyses.

Educators universally believe that active involvement facilitates learning. Yet the only empirical evidence from previous classroom research that supports this belief is indirect and equivocal. Specifically, several findings from the process-product studies (e.g. Anderson, Evertson, & Brophy, 1979; for a review see Rosenshine & Stevens, 1984) suggest that students who are actively involved in lessons make larger year-to-year gains in reading. A problem of interpretation of these findings arises, however, because students who were more involved may well have been more able, motivated, confident, or socially aggressive; growth in reading could have been related to one or more of these traits, instead of simply to involvement.

The finding from this study and its precursor (Anderson, Mason, & Shirey, 1984) that active turntaking facilitates learning and remembering permits an unambiguous interpretation. Remember that children did not bid for turns to read; turns were assigned in order from a randomly chosen starting point. Thus, the design assured that whether a student was a turntaker or nonturntaker at certain moments during a lesson was independent of the student's ability and other personal traits. Hence, the observed increments in what the student learned at those moments is attributable to active involvement per se.

Properties of the texts influenced performance in several ways. Page position had a pronounced influence on propositions recalled, short-answer responses, and oral reading errors. While the exact shape of the best fitting function varied depending upon the outcome measure, one common feature across measures was the decline in performance from the beginning to the end of stories. This result is roughly consistent with other data on the influence of the serial position on the learning and remembering of information in texts, except that an up turn in performance at the very end of the texts would be expected (Freebody & Anderson, 1986). Page position appeared several times in interactions with meaning versus surface emphasis and aspects of reading ability. The form of these interactions was for the teaching emphasis or ability index to have stronger effects at the beginning than at the end of stories. The overall decline in performance and the fact that major factors started to lose their potency over the course of stories suggests progressive failure of the processes required for comprehension. Another possibility to explain declining performance keys on the tests rather than the lessons; it could be that the children tended to get tired before completing the tests each day. Writing a whole story from memory is something these children had never done before. Composing their thoughts, spelling, and handwriting were clearly a challenge for many of them. The short-answer test, which followed the retelling of each story, also involved composing and writing a phrase or more for each item.

The importance of the information on pages strongly influenced recall of propositions. Furthermore, groups with high comprehension proved to be more sensitive to importance than groups lower in comprehension. Both of these results have precedents in previous research (Brown & Smiley, 1977; Meyer, Brandt, & Bluth, 1980) and are easily rationalized in terms of metacognitive theory. Less easy to understand is the fact that the density of information on a page was associated with a higher percentage of correct responses to short-answer questions and a lower percentage of oral reading errors.

Readability was notable for its failure to have any discernible effect on any measure or to enter into any significant interactions. This was true even though the readability of pages varied widely--from grade levels of 1.0 to 8.0 on the Fry (1977) scale. That readability proved impotent is consistent with Anderson and Davison's (1988) thesis that sentence length and word length (or frequency), the variables included in readability formulas, are not inherently very important for comprehension but are merely markers for the factors that are causally important, such as topic knowledge and topic interest. Notice that demands on topic knowledge and topic interest are operative mainly at the level of the whole story, whereas readability varied mainly at the level of the page in this study.

Lastly, several interesting findings involving reading ability emerged from the study. Especially interesting was the fact that the group measures, in particular the group fluency measure, had significant effects even though we entered them after the individual measures. Thus there is no escape from the conclusion that the group measures contain something more than or different from the individual measures.

It is really quite extraordinary that the average ability of a group of 6 to 10 children of sometimes varied abilities predicts the children's performance on outcome measures as well as or better than the abilities of the individual children. But this is only the most recent in a long string of studies (e.g. Barr & Dreeben, 1983; Juel, 1990; Weinstein, 1976) to give this result, so the finding must be taken seriously.

There are several reasons why group ability might predict outcomes better than individual ability. The first is that as teachers compose reading groups they may correct for errors of measurement in individual ability, based on their voluminous experience with the children's class performance. If a child is sick, distracted, or, on the other side, performs uncharacteristically well on a short, one-shot test, the teacher can override this faulty test score information when composing groups. The effect would be that the group measure is a more valid and reliable assessment of the ability of individual students than is the individual measure. This is a possibility that we have developed at length in another paper (Anderson, Wilkinson, Mason, Shirey, & Wilson, 1988). The second possible explanation is that the teacher may compose reading groups partly on the basis of noncognitive traits that are positively related

to lesson outcomes, such as effort, attention, and cooperation. The effect might be, in other words, for group "ability" to indirectly reflect other characteristics that produce superior performance. The third possible explanation is that the teacher may employ more effective instructional practices with high than with low groups. For instance, the pace of instruction appears to be faster in high groups (Allington, 1984; Barr & Dreeben, 1983). Thus, group "ability" may indirectly reflect differentially effective features of instruction that members of a group receive in common. A fourth possibility is that the culture of reading groups of different levels may differ with respect, for instance, to norms for paying attention (Eder & Felmlee, 1984). Group "ability" could reflect in part variations among groups in attention, and variations in attention could explain why group ability is more highly associated with outcomes than is individual ability.

The present study offers no strong grounds for choosing among possible explanations of the group-ability effect. Previous research does not either, although investigators have displayed a distinct preference for explaining the effect in terms of differences among groups in instructional practices or norms of behavior. The idea that group ability could represent additional information about children's traits used by the teacher when she composes groups has not been seriously considered. Yet, as far as we can see, this is as plausible an explanation as any other.

Another interesting finding from the present study is that, among the measures of reading ability, it was group fluency that had the strongest relationships with outcomes. This is surprising because the measures of fluency (individual as well as group) were psychometrically imperfect: The words were easy for the children and scores piled up at the top of the scale. In the current *zeitgeist*, comprehension is everything, so it will seem doubly surprising to some that an imperfect measure of fluency was a better predictor of a range of outcomes than a psychometrically strong standardized comprehension measure. Perhaps the explanation is that children in highly fluent groups experience a smooth and continuous reading that promotes the development of a connected representation of the story. In contrast, children in low-fluency groups experience frequent continuity-destroying pauses, false starts, mistakes, and backtracking that may undermine the formation of a coherent, connected story representation.

References

Aldrich, J., & Nelson, F. (1984). *Linear probability, logit, and probit models*. Beverly Hills, CA: Sage.

Allington, R. L. (1984). Context coverage and contextual reading in reading groups. *Journal of Reading Behavior, 16*, 85-96.

Anderson, L. M., Evertson, C. M., & Brophy, J. E. (1979). An experimental study of effective teaching in first-grade reading groups. *Elementary School Journal, 79*, 193-223.

Anderson, R. C., & Davison, A. (1988). Conceptual and empirical bases of readability formulas. In G. Green & A. Davison (Eds.), *Linguistic complexity and text comprehension: Readability issues reconsidered* (pp. 23-54). Hillsdale, NJ: Erlbaum.

Anderson, R. C., Mason, J. M., & Shirey, L. (1984). The reading group: An experimental investigation of a labyrinth. *Reading Research Quarterly, 20*, 6-38.

Anderson, R. C., Wilkinson, I., Mason, J. M., Shirey, L., & Wilson, P. T. (1988). Do errors on classroom reading tasks slow growth in reading? *Elementary School Journal, 88*, 267-280.

Au, K. H., & Mason, J. M. (1981). Social organizational factors in learning to read: The balance of rights hypothesis. *Reading Research Quarterly, 17*, 115-152.

Barr, R., & Dreeben, R., (1983). *How schools work*. Chicago: University of Chicago Press.

Brown, A. L., & Smiley, S. S. (1977). Rating the importance of structural units of prose passages: A problem of metacognitive development. *Child Development, 48*, 1-8.

Burstein, L. (1980). The analysis of multi-level data in educational research and evaluation. In D. C. Berliner (Ed.), *Review of research in education, vol. 8* (pp. 158-233). Washington, DC: American Educational Research Association.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Davidson, R. M. (1970). *The relationship between teachers' questions and pupils' responses during a directed reading activity and a directed reading-thinking activity*. Unpublished doctoral dissertation, University of Michigan, Ann Arbor.

Dreeben, R. (1984). First-grade reading groups: Their formation and change. In P. L. Peterson, L. C. Wilkinson, & M. Hallinan (Eds.), *The social context of instruction*. Orlando, FL: Academic Press.

Dreeben, R., & Barr, R. (1987). An organizational analysis of curriculum and instruction. In M. T. Hallinan (Ed.), *The social organization of the schools*. New York: Plenum Press.

Duffy, G. G. (1981). Teacher effectiveness research: Implications for the reading profession. In M. Kamil (Ed.), *Directions in reading research and instruction* (pp. 113-136). Washington, DC: National Reading Conference.

Eder, D., & Felmlee, D. (1984). The development of attention norms in ability groups. In P. Peterson, L. Wilkinson, & M. Hallinan (Eds.), *The social context of instruction* (pp. 189-208). Orlando, FL: Academic Press.

Felmlee, D., Eder, D., & Tsui, W. (1985). Peer influence on classroom attention. *Social Psychology Quarterly, 48,* 215-226.

Freebody, P., & Anderson, R. C. (1986). Serial position and rated importance in the recall of text. *Discourse Processes, 9,* 31-36.

Fry, E. (1977). Fry's readability graph: Clarification, validity, and extension to level 17. *Journal of Reading, 20,* 242-252.

Good, T. L., & Marshall, S. (1984). Do students learn more in heterogeneous or homogeneous groups? In P. L. Peterson, L. C. Wilkinson, & M. Hallinan (Eds.), *The social context of instruction*. Orlando, FL: Academic Press.

Guszak, F. J. (1967). Teacher questioning and reading. *Reading Teacher, 21,* 227-234.

Hallinan, M. (1984). Summary and implications. In P. L. Peterson, L. C. Wilkinson, & M. Hallinan (Eds.), *The social context of instruction*. Orlando, FL: Academic Press.

Hansen, J. (1981). The effects of inference training and practice on young children's comprehension. *Reading Research Quarterly, 18,* 391-417.

Hansen, J., & Pearson, P. D. (1982). *An instructional study: Improving the inferential comprehension of good and poor fourth-grade readers* (Tech. Rep. No. 235). Urbana-Champaign: University of Illinois, Center for the Study of Reading.

Hiebert, E. H. (1983). An examination of ability grouping for reading instruction. *Reading Research Quarterly, 18,* 231-255.

Hoffman, J. V., O'Neal, S. F., Kastler, L. A., Clements, R. O., Segel, K. W., & Nash, M. F. (1984). Guided oral reading and miscue focused verbal feedback in second-grade classrooms. *Reading Research Quarterly, 19,* 367-384.

Johnson, D. W. (1981). Student-student interaction: The neglected variable in education. *Educational Researcher, 10,* 5-10.

Jose, P. E. & Brewer, W. F. (in press). Early grade children's liking of script and suspense story structures. *Journal of Reading Behavior.*

Juel, C. (1990). Effects of reading group assignment on reading development in first and second grade. *Journal of Reading Behavior, 22,* 233-254.

Martin, J., Veldman, D. J., & Anderson, L. M. (1980). Within-class relationships between student achievement and teacher behaviors. *American Educational Research Association Journal, 17,* 479-490.

Mason, J. M. (1976). Overgeneralization in learning to read. *Journal of Reading Behavior, 8,* 173-182.

Mason, J. M. (1984). A question about reading comprehension instruction. In G. G. Duffy, L. R. Roehler, & J. Mason (Eds.), *Comprehension instruction: Perspectives and suggestions* (pp. 39-56). New York: Longman.

McKelvey, R. D., & Zavoina, W. (1975). A statistical model for the analysis of ordinal level dependent variables. *Journal of Mathematical Sociology, 4,* 103-120.

Meyer, B. J. F., Brandt, D. M., & Bluth, G. J. (1980). Use of top-level structure in text: Key for reading comprehension of ninth-grade students. *Reading Research Quarterly, 16,* 72-103.

O'Flahavan, J. F., Hartman, D. K., & Pearson, P. D. (1989). *Teacher questioning and feedback practices after the cognitive revolution: Replication and extension of Guszak's (1967) study* (Tech. Rep. No. 461). Urbana-Champaign: University of Illinois, Center for the Study of Reading.

Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction, 1,* 117-175.

Pedhazur, E. J. (1982). *Multiple regression in behavioral research*. New York: Holt, Rinehart & Winston.

Pehrsson, R. S. V. (1974). The effects of teacher interference during the process of reading or how much of a helper is Mr. Gelper? *Journal of Reading, 17,* 617-621.

Persell, C. (1977). *Education and inequality: The roots and results of stratification in America's schools*. New York: Free Press.

Petre, R. M. (1970). *Quantity, quality and variety of pupil responses during an open-communication structured group directed reading-thinking activity and a closed-communication structured group directed reading activity*. Unpublished doctoral dissertation, University of Delaware, Newark.

Raphael, T. E., & Wonnacott, C. A. (1985). Heightening fourth-grade students' sensitivity to sources of information for answering comprehension questions. *Reading Research Quarterly, 20,* 282-296.

Raudenbush, S. W., & Bryk, A. S. (1988). Methodological advances in analyzing the effects of schools and classrooms on student learning. In E. Z. Rothkopf (Ed.), *Review of research in education: Vol. 15* (pp. 423-475). Washington, DC: American Educational Research Association.

Redfield, D. L., & Rousseau, E. W. (1981). A meta-analysis of experimental research on teacher questioning behavior. *Review of Educational Research, 51,* 237-245.

Rosenshine, B., & Stevens, R. (1984). Classroom instruction in reading. In P. D. Pearson (Ed.), *Handbook of reading research* (pp. 745-798). New York: Longman.

SAS Institute (1986). *SUGI Supplemental Library User's Guide* (Version 5 ed.). Cary, NC: SAS Institute Inc.

Slavin, R. E. (1987). Ability grouping and student achievement in elementary schools: A best-evidence synthesis. *Review of Educational Research, 57,* 293-336.

Sorensen, A. B., & Hallinan, H. (1984). Effects of race on assignment to ability groups. In P. L. Peterson, L. C. Wilkinson, & M. Hallinan (Eds.), *The social context of instruction*. Orlando, FL: Academic Press.

SPSS (1988). *SPSS^x User's Guide* (3rd ed.). Chicago: SPSS Inc.

SRA assessment survey: Achievement series (1971). *Primary edition. Reading-Form E/Primary II*. Chicago: Science Research Associates.

Tierney, R. J., & Cunningham, J. W. (1984). Research on teaching reading comprehension. In P. D. Pearson (Ed.), *Handbook of reading research* (pp. 609-655). New York: Longman.

van den Broek, P. (1990). Causal inferences and the comprehension of narrative texts. In A. Graesser & G. Bower (Eds.), *Psychology of learning and motivation: Inferences and text comprehension.* New York: Academic Press.

Weinstein, R. S. (1976). Reading group membership in first grade: Teacher behaviors and pupil experience over time. *Journal of Educational Psychology, 68,* 103-116.

Author Note

The authors are grateful to Pose Lamb, Larry Shirey, and Paul Wilson for their assistance in completing this research and to Jere Brophy for his valuable comments on a draft of this report.

Table 1

Assignment of Treatments and Stories to Classes According to Day

Class	Day 1	Day 2	Day 3	Day 4	Day 5
1	St A--ME	St B--ME	Tests	St C--SE	St D--SE
2	St A--SE	St B--SE	Tests	St C--ME	St D--ME
3	St C--ME	St D--ME	Tests	St A--SE	St B--SE
4	St C--SE	St D--SE	Tests	St A--ME	St B--ME
5	St B--ME	St A--ME	Tests	St D--SE	St C--SE
6	St B--SE	St A--SE	Tests	St D--ME	St C--ME

Table 2

Description of Variables

Variable	Measurement Scale/Method	Meaning of High Positive Score	*M*	*SD*
Dependent Variables				
Propositions[a]	Percentage	Good Recall	14.00	18.30
Important Elements[a]	Percentage	Good Recall	35.30	21.40
Short Answers	1, 0	Correct Response	.63	.48
Interest	Ranking	Very Interesting	2.52	1.12
Oral Reading Errors[a]	Percentage	Many	1.40	2.70
Lesson Time	Minutes	Long	14.00	3.22
Independent Variables				
Meaning vs. Surface	+1, -1	Meaning emphasis	0.00	1.00
Turntaking	+1, -1	Active Turn	-.75	.66
Individual Comprehension	National Stanine	High	6.64	1.95
Individual Fluency[a]	Local Stanine	Fluent	5.00	2.00
Turntaker's Comprehension	National Stanine	High	6.64	1.95
Turntaker's Fluency[a]	Local Stanine	Fluent	5.00	2.00
Group Comprehension	Group Mean	High	6.62	1.29
Group Fluency[a]	Group Mean	Fluent	4.99	1.18
Group Heterogeneity in Comprehension	Standard Deviation	Diverse	1.33	.73
Group Heterogeneity in Fluency[a]	Standard Deviation	Diverse	1.52	.72
Gender	+1, -1	Girl	.05	1.00
Ethnicity	+1, -1	Black	-.49	.88
Importance of Page	Proportion	Important	.10	.06
Readability of Page	Grade Level	Difficult	2.78	1.43
Density of Page	Proportion	Dense	.08	.01
Position of Page	Serial Position	End	6.50	3.49

[a]Normalizing transformation applied before analysis.

Table 3

Means of Dependent Variables by Meaning Versus Surface Emphasis

Variable	Scale	Meaning	Surface
Recall of propositions	Percentage	14.83	13.08
Short answers	Probability	.65	.62
Recall of important elements	Percentage	37.38	33.24
Reading errors	Percentage	1.16	1.69
Story interest	Ranking	2.72	2.31
Lesson time	Minutes	12.81	15.19

Table 4

Means of Dependent Variables by Turntaker Versus Nonturntaker

Variable	Scale	Turntaker	Nonturntaker
Recall of propositions	Percentage	15.48	13.74
Short answers	Probability	.67	.63

Table 5

Subject-Level Analysis of Recall of Propositions

Variable	b-weights Entry	b-weights Final	Percent Variance	*F*
Between-Subjects				
Subject's comprehension	.05	.03	11.76	20.47*
Subject's fluency	.03	.01	3.32	5.79
Group fluency	.07	.08	4.38	7.62*
Group comprehension	-.02	-.02	.15	.26
Constant/residual		-.01	80.39	
Within-Subjects				
Meaning vs. Surface	.03	.16	.43	36.74*
Importance of page	1.23	-.54	1.96	167.83*
Page position	-.05	-.03	9.12	782.93*
(Page position)^2	.002	.002	.21	18.11*
Active turn	.03	.03	.19	16.14*
Meaning vs. Surface x Page position	-.004	-.004	.05	4.55
Meaning vs. Surface x Group comprehension	-.01	-.01	.12	10.13*
Importance of page x Group comprehension	.26	.24	.13	11.42*
Page position x Subject's comprehension	-.005	-.003	.35	29.79*
Page position x Group fluency	-.005	-.005	.13	10.90*
Other page and story factors			11.04	20.16*
Constant/residual		.14	76.28	

* For Between-Subjects factors, critical F (1,140) = 6.82, $p < .01$; for Within-Subjects factors, critical F (1,6548) = 6.63 $p < .01$; for other page and story factors, critical F (47,6548) = 1.72, $p < .01$.

Table 6

Group-Level Analysis of Recall of Propositions

Variable	b-weights Entry	b-weights Final	Percent Variance	*F*
Between-Groups				
Group fluency	.10	.09	40.45	7.51
Group comprehension	.01	.01	.29	.05
Constant/residual		.01	59.26	
Within-Groups				
Meaning vs. Surface	.03	.14	1.27	28.77*
Importance of page	1.22	-.54	5.61	127.34*
Page position	-.05	-.04	26.92	611.10*
(Page position)^2	.002	.002	.60	13.71*
Meaning vs. Surface x Page position	-.003	-.003	.14	3.09
Meaning vs. Surface x Group comprehension	-.01	-.01	.26	5.94
Importance of page x Group comprehension	.25	.24	.38	8.72*
Page position x Group fluency	-.01	-.01	.95	21.63*
Other page and story factors			32.54	15.62*
Constant/residual		.19	31.32	

* For Between-Groups factors, critical F (1,11) = 9.65, $p < .01$; for Within-Groups factors, critical F (1,711) = 6.68, $p < .01$; for other page and story factors, critical F (47,711) = 1.73, $p < .01$.

Table 7

Subject-Level Analysis of Short Answers

Variable	b-weights Entry	b-weights Final	Percent Variance	X^2
Between-Subjects				
Subject's comprehension	.58	.48	27.73	30.99*
Subject's fluency	.33	-.01	6.27	8.30*
Group fluency	1.42	1.79	24.17	25.33*
Group comprehension	-.36	-.36	3.85	1.06
Constant/residual		-7.72	37.98	
Within-Subjects				
Meaning vs. Surface	.08	.23	.17	8.64*
Density of page	16.80	16.14	1.24	61.27*
Page position	-.06	.13	1.24	62.82*
(Page position)^2	-.01	-.01	.20	13.13*
Active turn	.09	.33	.10	5.92
Turntaker's fluency	-.04	-.07	.06	2.66
Meaning vs. Surface x Page position	-.02	-.02	.13	7.26*
Active turn x Turntakers's fluency	-.05	-.05	.09	6.71*
Page position x Group comprehension	-.01	-.01	.15	6.10
Other page and story factors			23.76	1398.40*
Constant/residual		-3.75	72.86	

* Critical X^2 (1) = 6.63, $p < .01$; for other page and story factors, critical X^2 (47) = 27.44, $p < .01$.

Table 8

Group-Level Analysis of Short Answers

Variable	b-weights Entry	b-weights Final	Percent Variance	*F*
Between-Groups				
Group fluency	.09	.08	70.00	26.79*
Group comprehension	.02	.02	1.26	.48
Constant/residual		.12	28.74	
Within-Groups				
Meaning vs. Surface	.02	.04	.41	8.16*
Density of page	3.22	3.07	2.59	51.29*
Page position	-.01	.03	2.55	50.38*
(Page position)^2	-.002	-.002	.70	13.94*
Meaning vs. Surface x Page position	-.004	-.004	.25	4.91
Page position x Group comprehension	-.002	-.002	.14	2.69
Other story and page factors			57.45	24.11
Constant/residual		-.21	35.92	

* For Between-Groups factors, critical F (1,11) = 9.65, $p < .01$; for Within-Groups factors, critical F (1,711) = 6.68, $p < .01$; for other story and page factors, critical F (47,711) = 1.73, $p < .01$.

Table 9

Subject-Level Analysis of Recall of Important Elements

Variable	b-weights Entry	b-weights Final	Percent Variance	*F*
Between-Subjects				
Subject's comprehension	.09	.06	17.85	35.45*
Subject's fluency	.05	.01	4.10	8.14*
Group fluency	.13	.14	7.52	14.93*
Group comprehension	-.01	-.01	.04	.07
Constant/residual		.21	70.49	
Within-Subjects				
Story	a	a	11.11	17.58*
Meaning vs. Surface	.05	.20	3.02	14.33*
Meaning vs. Surface x Group Comprehension	-.02	-.02	.99	4.72
Constant/residual		.20	84.88	

* For Between-Subjects factors, critical F (1,140) = 6.82, $p < .01$; for Within-Subjects factors, critical F (1,403) = 6.70, $p < .01$, except for Story, where critical F (3,403) = 3.83.

[a] Three orthogonal contrasts omitted.

Table 10

Group-Level Analysis of Recall of Important Elements

Variable	b-weights Entry	b-weights Final	Percent Variance	*F*
Between-Groups				
Group fluency	.18	.15	57.64	15.54*
Group comprehension	.04	.04	1.56	.42
Constant/residual		.22	40.80	
Within-Groups				
Story	a	a	38.06	9.78*
Meaning vs. Surface	.05	.19	9.77	7.53*
Meaning vs. Surface x Group Comprehension	-.02	-.02	2.90	2.24
Constant/residual		.22	49.27	

* For Between-Groups factors, critical F (1,11) = 9.65, $p < .01$; for Within-Groups factors, critical F (1,38) = 7.35, $p < .01$, except for Story where critical F (3,38) = 4.34.

[a] Three orthogonal contrasts omitted.

Table 11

Analysis of Oral Reading Errors

Variable	b-weights Entry	b-weights Final	Percent Variance	*F*
Between-Turntakers				
Turntaker's fluency	-.07	-.05	26.86	59.60*
Turntaker's comprehension	-.05	-.05	10.49	23.26*
Turntaker's race	-.003	-.003	.01	.02
Constant/residual		-1.49	62.65	
Within-Turntakers				
Meaning vs. Surface	-.04	.03	2.36	17.11*
Density of page	-2.60	-2.51	1.94	14.05*
Page position	.002	.02	.08	.55
(Page position)^2	-.001	-.001	.30	2.14
Meaning vs. Surface x Page position	-.001	-.03	.04	.26
Meaning vs. Surface x (Page position)^2	.003	.003	1.33	9.62*
Meaning vs. Surface x Turntaker's race	-.03	-.03	1.52	11.04*
Other story and page factors			13.37	2.05*
Constant/residual		-1.35	79.07	

* For Between-Turntakers factors, critical F (1,139) = 6.82, $p < .01$; for Within-Turntakers factors, critical F (1,573) = 6.69, $p < .01$; for other story and page factors, critical F (47,573) = 1.73, $p < .01$.

Table 12

Subject-Level Analysis of Story Interest

Variable	b-weights		Percent Variance	X^2
	Entry	Final		
Within-Subjects				
Story	a	a	16.09	93.85*
Meaning vs. Surface	.33	.33	2.58	18.06*
Constant/residual	b	b	81.11	

* For Story, critical X^2 (3) = 11.34, $p < .01$; for Meaning vs. Surface, critical X^2 (1) = 6.63, $p < .01$.

[a] Three orthogonal contrasts omitted.

[b] Three intercepts omitted.

Table 13

Group-Level Analysis of Story Interest

Variable	b-weights		Percent Variance	*F*
	Entry	Final		
Within-Groups				
Story	a	a	43.35	11.59*
Meaning vs. Surface	.20	.20	9.30	7.46*
Constant/residual		2.52	47.35	

* For Story, critical F (3,38) = 4.34 $p < .01$; for Meaning vs. Surface, critical F (1,38) = 7.35, $p < .01$.

[a] Orthogonal contrasts omitted.

Table 14

Analysis of Lesson Time

Variable	b-weights		Percent Variance	*F*
	Entry	Final		
Between-Groups				
Group fluency	-1.58	-1.58	60.03	16.52*
Constant/residual		21.87	39.97	
Within-Groups				
Meaning vs. Surface	-1.19	-1.19	30.99	17.06*
Constant/residual		14.00	69.01	

* For Between-Groups factors, critical F (1,11) = 9.65, $p < .01$; for Within-Groups factors, critical F (1,38) = 7.35, $p < .01$.

Figure Captions

Figure 1. Recall of propositions as a function of meaning versus surface emphasis and group comprehension.

Figure 2. Recall of propositions as a function of average group comprehension and importance of page.

Figure 3. Recall of propositions as a function of individual comprehension and page position.

Figure 4. Recall of propositions as a function of average group fluency and page position.

Figure 5. Responses to short answer questions as a function of meaning versus surface emphasis and page position.

Figure 6. Responses to short answer questions as a function of turntaking and the turntaker's fluency.

Figure 1

Recall of Propositions as a Function of Meaning Versus Surface Emphasis and Group Comprehension.

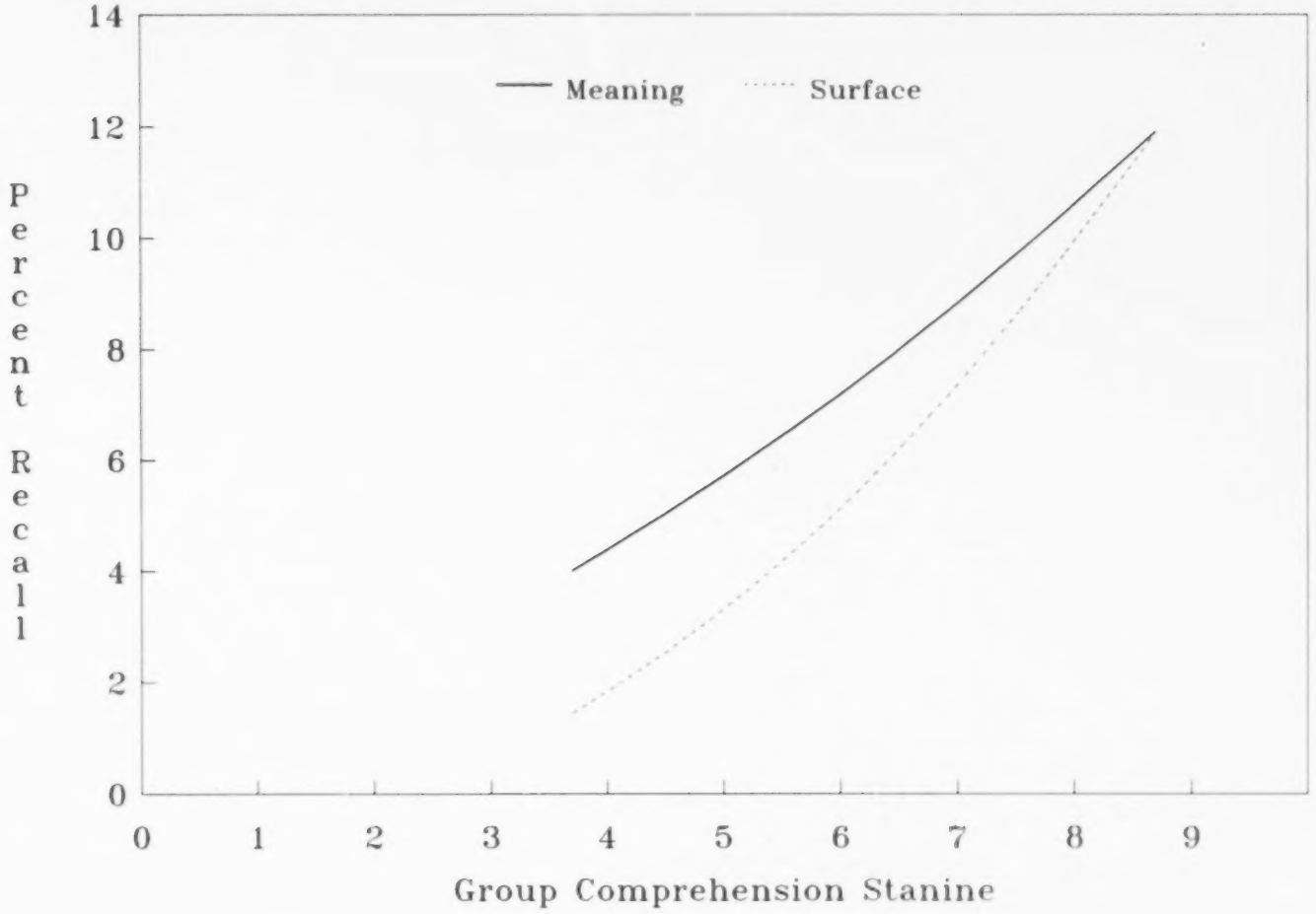

Figure 2

Recall of propositions as a function of average group comprehension and importance of page.

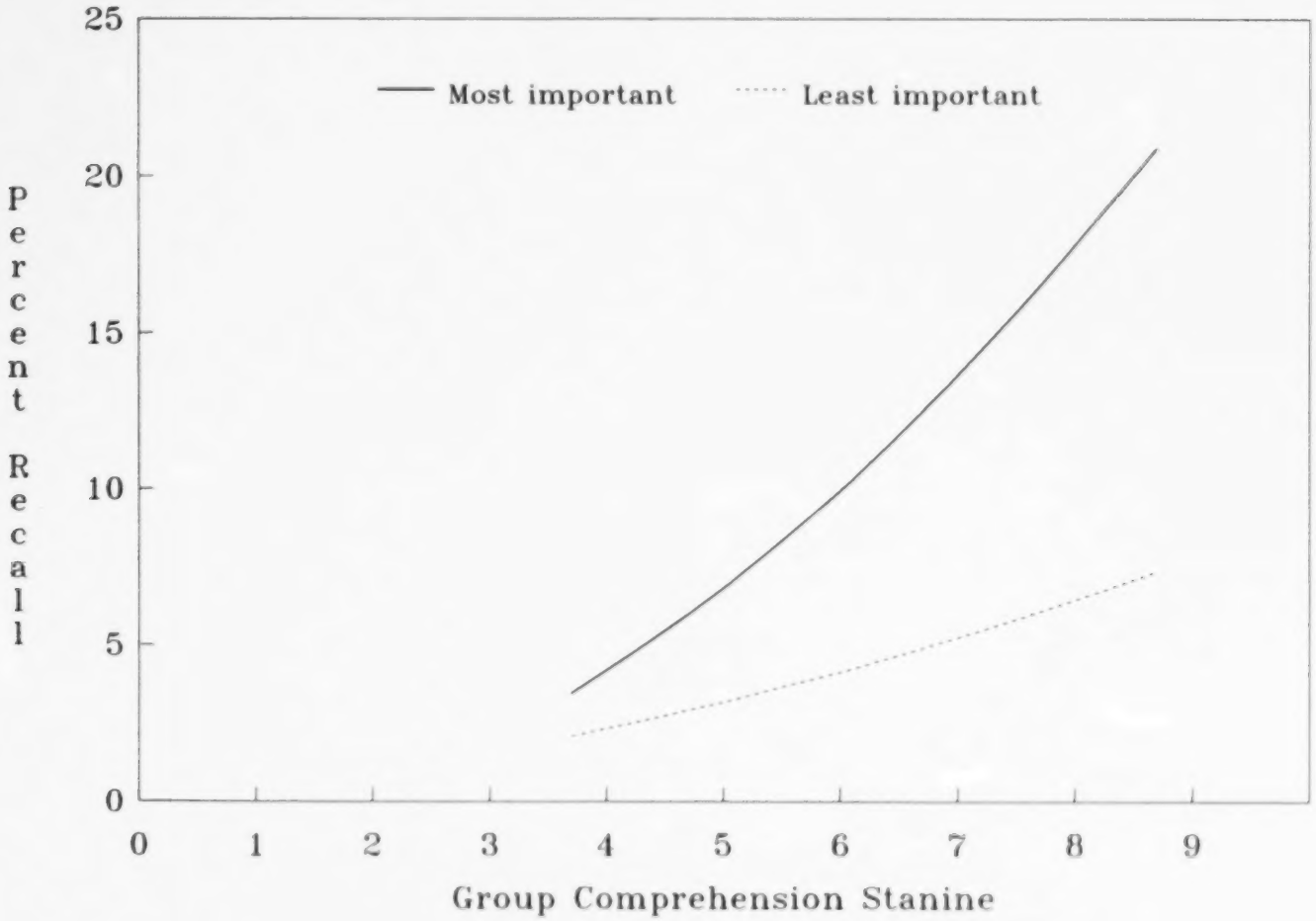

Figure 3

Recall of Propositions as a Function of Individual Comprehension and Page Position.

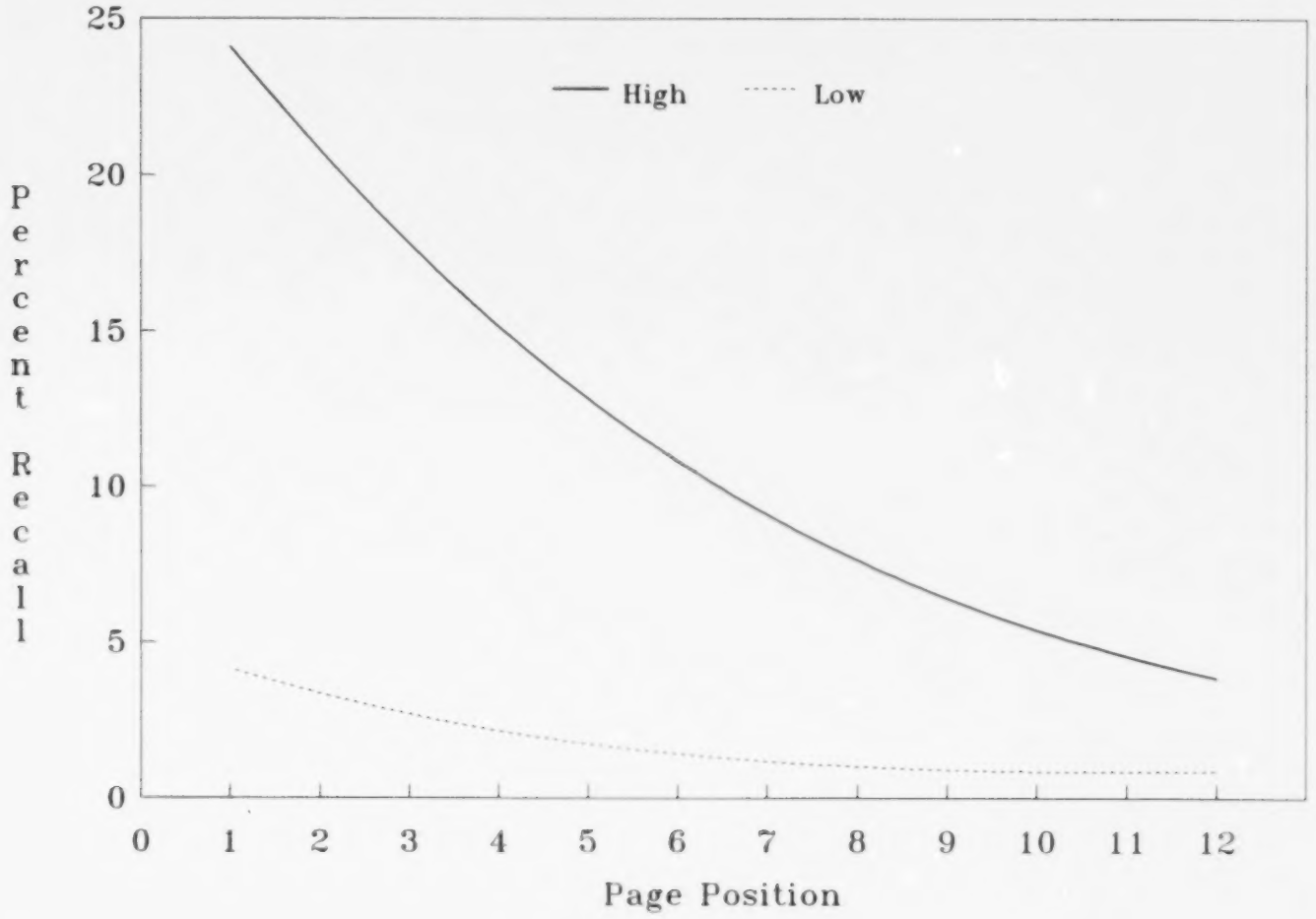

Figure 4

Recall of Propositions as a Function of Average Group Fluency and Page Position.

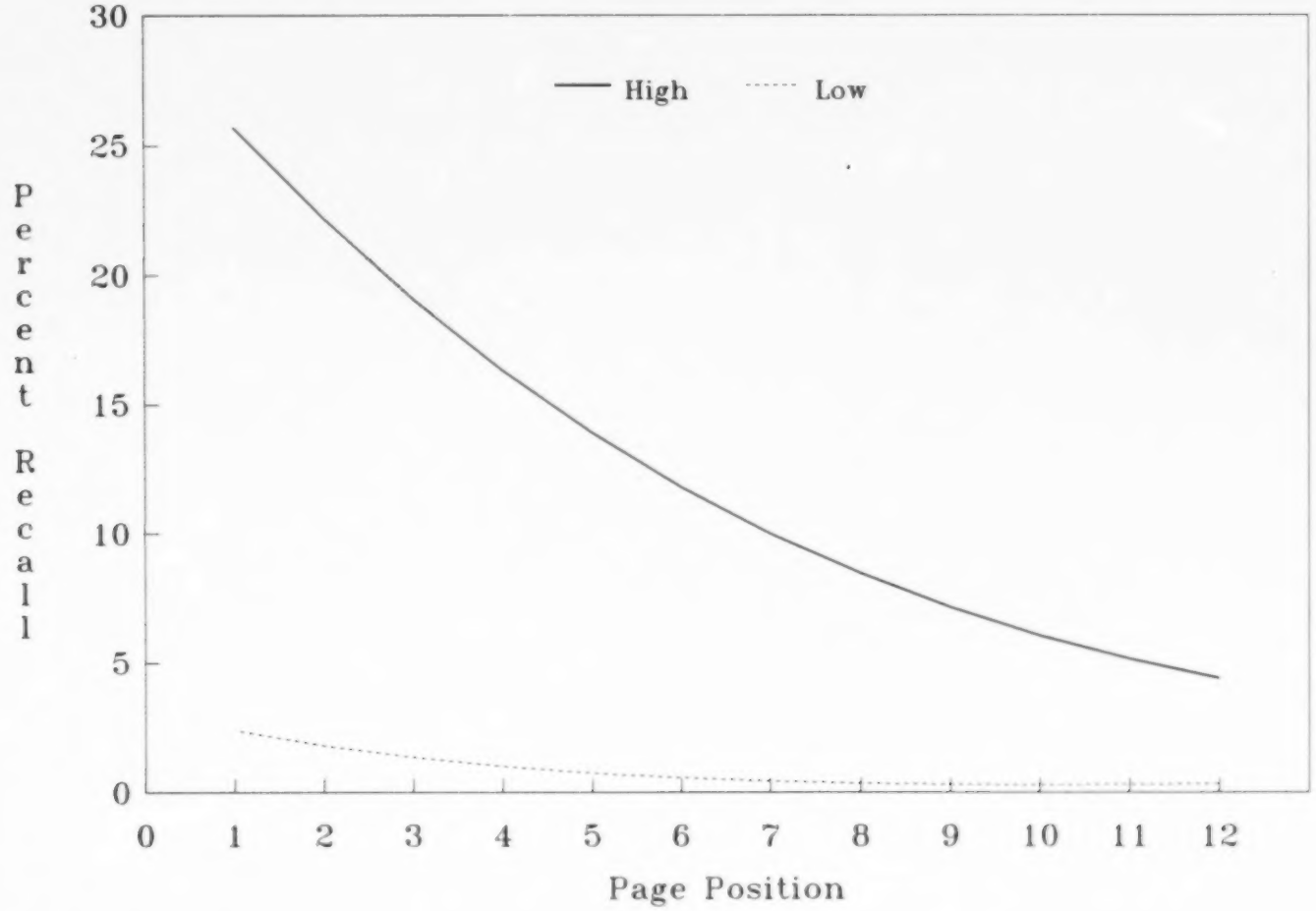

Figure 5

Responses to Short Answer Questions as a Function of Meaning Versus Surface Emphasis and Page Position.

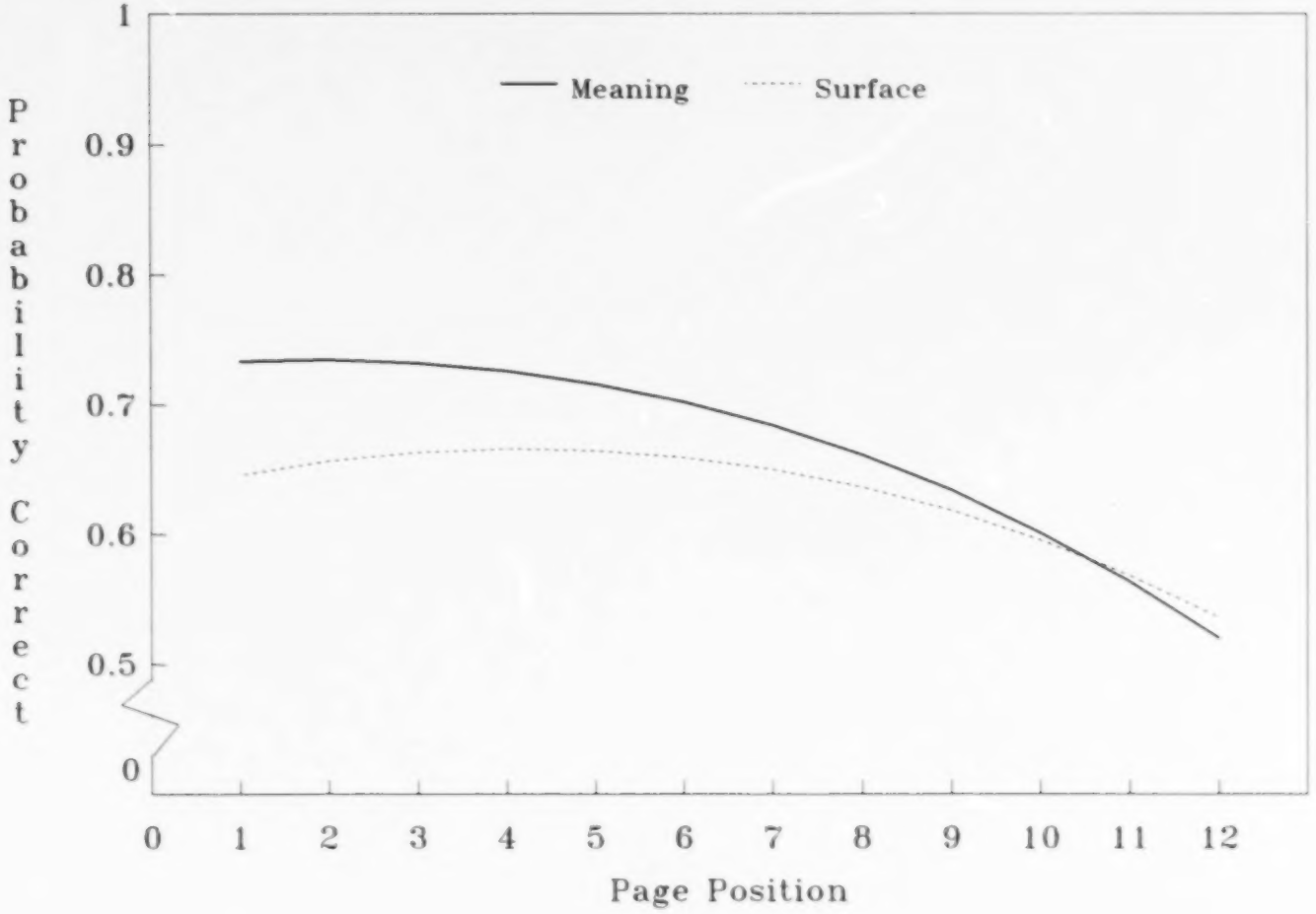

Figure 6

Responses to Short Answer Questions as a Function of Turntaking and the Turntaker's Fluency.